21742

NATIONAL LIBRARY
OTTAWA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR CATHY D. AITKENS

TITLE OF THESIS EFFECTS OF TEACHER EXPECTATIONS ON STUDENT ACHIEVEMENT IN SWIMMING

UNIVERSITY OF ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED M.A

YEAR THIS DEGREE GRANTED 1974

Permission is hereby granted to THE NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed) Cathy Aitkens

PERMANENT ADDRESS:

61 LUCIEN DR.

DARTMOUTH

NOVA SCOTIA

DATED OCTOBER 4 1974

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

EFFECTS OF TEACHER EXPECTATIONS ON STUDENT ACHIEVEMENT IN SWIMMING

BY

© CATHY AITKENS

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

DEPARTMENT OF PHYSICAL EDUCATION

EDMONTON, ALBERTA

FALL, 1974

Richard W. Scott

Scott

Date ... October ... 1971 ...

ABSTRACT

The central purpose of this study was to investigate the effectiveness of a methodology used to study the Pygmalion effect in a physical education environment. An attempt was made to adapt a methodology, which had been used successfully in a classroom, to the swimming pool. Three sub-problems were examined: (1) the experimental manipulation of teacher expectations, (2) the effects of teacher expectations on teacher behavior, and (3) the effects of teacher expectations on the student's self-rating of and improvement in swimming.

The results indicated that there was no reason to believe that teacher expectations had been successfully manipulated and therefore, no conclusions could be drawn related to the effects of teacher expectations on teacher behavior and student behavior.

Subsequent examination of the methodology resulted in recommendations for changes which may lead to a more effective approach for studying this problem.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the members of my supervising committee. Their involvement and encouragement created the environment for this positive learning experience. Special thanks to Dr. Harvey Scott.

TABLE OF CONTENTS

	PAGE
CHAPTER I - STATEMENT OF THE PROBLEM	1
1. Introduction	1
2. Justification of the Study	2
3. The Problem	3
4. Delimitations	4
5. Definition of Terms	5
CHAPTER II - REVIEW OF THE LITERATURE	7
1. Teacher Expectancy Effect	7
2. Prophecy Communication	17
3. Teacher Expectancy and Student Self-Concept	19
CHAPTER III - METHODS AND PROCEDURES	30
1. Selection of Sample	30
2. Selection and Construction of Instruments	30
3. Collection of Data	32
4. Analysis of Data	34
CHAPTER IV - RESULTS AND DISCUSSION	39
1. Experimental Manipulation of Teacher Expectations	39
2. Effects of Teacher Expectations on Teacher Behavior	42
3. Effects of Teacher Expectations on Student Self-Rating and Skill Improvement	47

PAGE

4. Discussion of the Results 41

5. Methodological Implications 47

CHAPTER V - SUMMARY AND CONCLUSIONS 48

1. Summary 48

2. Conclusions 48

3. Recommendations 48

BIBLIOGRAPHY 49

APPENDIX A 55

APPENDIX B 61

LIST OF TABLES

TABLE		PAGE
1.	Mean Differences Between the Teacher Ratings and the Expert Ratings	37
2.	Frequencies of Specific Teacher Behaviors	39
3.	Product-Moment Correlation Coefficients Between Teacher Ratings and Other Ratings of Student Swimming Ability	40
4.	Product-Moment Correlations Between the Self-Ratings and the Expert Ratings of Student Swimming Ability	42
5.	Comparison of the Mean Differences in Self-Ratings and the Expert Ratings	44

CHAPTER I

STATEMENT OF THE PROBLEM

I. Introduction

An area of physical education which has virtually escaped research is the effect of teacher expectations on the performance of students. "If it is true that some of the variance in achievement is produced by the expectations for performance held by the teachers and students, educators are somehow defeating their own ends." (Hoffman and Cohen, 1972:2)

Related classroom research of this phenomenon does evidence to support its occurrence in some circumstances. Robert Rosenthal has carried out extensive studies (1968, 1969, 1968, 1971, 1973) in an attempt to evaluate the direction and intensity of teacher expectancy effects, i.e. the Pygmalion effect, under various conditions. Although no conclusive evidence is yet available, Rosenthal and other researchers have begun to define the complex nature of the variables involved and to solve some of the methodological difficulties encountered while conducting such research.

The need for further serious research in this area becomes evident when one looks at the possible negative effects which may occur as a result of teacher expectations. Kenneth Clark (1963), in his study of racially disadvantaged children, does a tragic illustration of this:

> If a child scores low on an intelligence test because he cannot read and then is not taught to read because he has a low score, then such a child is being imprisoned in an iron circle and becomes the victim of an educational self-fulfilling prophecy. (1963:190)

Perhaps classroom teachers are not the only ones who unconsciously perpetuate this phenomenon. Is it not possible that physical education teachers as well, interact with their students in a way which will affect the fulfillment of the expectations they hold for student performance?

It is the intent of this study to investigate this possibility.

2. JUSTIFICATION OF THE STUDY

With the exception of one study, the research investigating the effects of teacher expectations on student achievement has been conducted in classroom settings. This study therefore, can be justified for it is an attempt to develop an effective methodology for exploring this phenomenon in a physical education environment.

3. THE PROBLEM

The main focus of this study was to explore the effectiveness of the methodology which had been designed to investigate the Pygmalion effect in physical education. In order to determine this effectiveness, a case study was conducted in which three sub-problems were chosen for investigation. The first was the attempt to experimentally induce the phenomenon of the self-fulfilling prophecy in an instructional swimming situation and to observe its effects. The second was to explore some aspects of the teacher behaviors of feedback, reward, correction and prohibition as potential determinants of the self-fulfilling prophecy. The third was to investigate what effects teacher expectations may have on the self-rating of swimming ability and the skill improvement of the students. In order to structure this investi-

sation more specifically, the following analyses and comparisons were made:

A. A comparison of the expert ratings and the teacher ratings of the swimming ability of the "special potential" group and the control group.

B. A comparison of the frequencies of the feedback, reward, correction and prohibiting behaviors exhibited by the teacher to the "special potential" group and the control group.

C. An analysis of the similarities between teacher expectations and ratings of the students, students' perceptions of teacher expectations and ratings of them, and students self expectations and ratings.

D. A comparison of the expectations, self-ratings and perceived teacher ratings of swimming ability of the "special potential" group and the control group.

E. A comparison of the improvement in swimming skills of the "special potential" group and the control group.

Hypotheses: Based on the above comparisons, the following hypotheses were formulated:

1. The teacher's ratings of swimming ability will be significantly greater for the "special potential" group than for the control group.

2. The difference between the teacher's ratings and the expert's post ratings will be significantly greater for the "special potential" group than for the control group.

3. The frequencies of feedback and reward behaviors exhibited by the teacher will be significantly greater for the "special potential" group than for the control group.

4. The frequencies of correction and prohibitive behaviors exhibited by the teacher will be significantly greater for the control group than for the "special potential" group.

5. The correlation between the expert ratings and the teacher rating will be significantly greater for the control group than for the "special potential" group.

5. The correlation between the students' self-ratings, from the reference point of the teacher, and the teacher ratings will be significantly greater for the "special potential" group than for the control group.

6. There will be no significant difference in the correlation between the post self-ratings and the teacher ratings for the control group and the "special potential" group.

7. The correlation between the self-ratings and the expert ratings will be significantly greater for the control group than for the "special potential" group.

8. The correlation between the self-ratings, from the reference point of the teacher, and the self-ratings will be significantly greater for the "special potential" group than for the control group.

9. The correlation between the potential self-ratings and the post self-ratings will be significantly greater for the "special potential" group than for the control group.

10. The post self-ratings of swimming ability will be significantly greater for the "special potential" group than for the control group.

11. The post self-ratings, from the reference point of the teacher, will be significantly greater for the "special potential" group than for the control group.

12. The difference between the potential self-ratings and the post self-ratings will be significantly greater for the "special potential" group than for the control group.

13. The improvement in swimming ability will be significantly greater for the "special potential" group than for the control group.

4. DELIMITATIONS

1. The sampling of subjects was limited to 10 boys and girls, ages 9 to 10 years, enrolled in the instructional swimming program at the University of Alberta and their teacher.

2. The subjects were limited to those enrolled for swimming instruction at the Junior Red Cross level.

5. DEFINITION OF TERMS

Self-Fulfilling Prophecy. The self-fulfilling prophecy refers to the phenomenon in which one person's expectations for another's behavior can come to serve as a partial determinant of that behavior.

Pygmalion Effect. The Pygmalion effect is a synonymous term for the self-fulfilling prophecy. (Rosenthal, 1969)

"Special Potential" Students. The 'special potential' students were the students who were arbitrarily designated as those who possessed a special potential for improvement in swimming ability.

Student Self-rating of Swimming Ability. The student self-rating of swimming ability was the rating which the students gave to themselves based on their perception of how well they could swim.

Student Self-rating of Potential Swimming Ability. The student self-rating of potential swimming ability was the rating which the students gave to themselves based on their perception of how well they felt they would be able to swim at the end of the course.

Student Self-rating of Swimming Ability from the Reference Point of the Teacher. The student self-rating of swimming ability from the reference point of the teacher was the rating which the students gave to themselves based on how well they felt their teacher would rate them.

Teacher Rating of Student's Swimming Ability. The teacher rating of student's swimming ability was the rating the teacher gave to

each student based on how well he felt the student could swim.

Expert Rating of Student Swimming Ability. The expert rating of student's swimming ability was the rating a panel of experts gave to the student based on how well the student could actually swim.

CHAPTER II

REVIEW OF THE LITERATURE

The review of the literature has been divided into three main areas; the first section dealing with the teacher expectancy effect, the second dealing with the communication of a prophecy, and the third dealing with the effect of teacher expectations on the self concept of the student.

1. TEACHER EXPECTANCY EFFECT

Although research on the self-fulfilling prophecy began as early as the mid-1920's, with the Hawthorne effect (Roethlisberger and Dickson, 1939), it has only been during the last decade that the educational implications of this phenomenon have received major attention in research.

Just what are the effects of teacher's expectations on the achievement of their pupils? The major attempt to answer this question stems from the work of Robert Rosenthal and Lenore Jacobson and their Oak School experiment (1968). This initial study involved the children of an elementary school in a lower socio-economic district. A non-verbal test of intelligence was administered in a few classes in each of grades one to six. This intelligence test was disguised as a test to determine late "intellectual bloomers". The teachers were informed that they would be given the names of those students who would show the greatest intellectual gains over the coming year. Actually, about twenty percent of the children who had taken the test were randomly selected and designated as "intellectual bloomers". It was the names

of these children which were given to the teachers. At the end of the school year, eight months later, the intelligence test was again administered to all of the children who had been pretested.

This study was designed to test the hypothesis that those children from whom the teacher expected greater intellectual achievement would actually show such achievement. Rosenthal and Jacobson felt that merely telling the teacher that some of her students were "intellectual bloomers" would result in changes in the pupils' IQs. The data revealed that the children in the experimental group showed a greater gain of four points in total IQ and seven points in Reasoning IQ when compared with the children in the control group. With the Reasoning IQ, the experimental girls showed an expectancy advantage while the experimental boys showed an expectancy disadvantage. It was also found that the differences shown in the first two grades were more permanent than those in the remaining grades. Children from the slowest classes as well as those from the most advanced classes revealed equal expectancy benifits. Substantial interactions between the sex of a student, grade level and IQ subtest were revealed in subsequent analysis of the data.

Two partial replications (Rosenthal and Evans, 1968; Conn, Edwards, Rosenthal and Crowne, 1967) of this early study were conducted in an attempt to verify its results. Rosenthal and Evans found a dramatic difference in their study of two schools in a small middle class midwestern town. Favorable expectancy effects were revealed in boys' Reasoning IQs only, with the girls showing an expectancy dis-

advantage in Reasoning IQ. No significant results were obtained for either Verbal IQ or total IQ.

The setting for the Conn, Edwards, Rosenthal and Crowne study was in an eastern upper middle class suburban school. The procedure was essentially the same as the Oak School experiment except that the pretesting was done at the beginning of the second semester. The teachers, therefore, had the first semester to form expectations of their pupils. The retesting was conducted four months later, at the end of the semester. The boys showed the greatest effects with a mean increase of eight points in Verbal IQ. Other combinations of sex and IQ subtest showed significant expectancy advantages. When compared with the Rosenthal and Jacobson study, all of the children of the study by Conn et al. showed a greater IQ increase. One year later, however, a retest revealed opposite results with the control children showing greater gains than the experimental children. The results of these quasi-replications of the Oak School experiment reveal the complex nature of the relevant variables and the instability of the effects.

Other researchers have attempted to demonstrate expectancy effects in a variety of situations. Biegen (Rosenthal, 1968) and Flowers (1966) manipulated teacher expectations for whole classes and not just specific children within classes. In both cases, the classes expected by the teacher to show greater gains in IQ, produced these results. Burnham (1968) worked with swimming instructors at a summer camp who were teaching preadolescents how to swim. He found that it was the older boys (10-14 years) and the younger girls (7-9 years)

who were significantly influenced by the expectations of their instructors. An important point about this study was that the expectancy results were shown after only a two-week period of instruction. Also, the results revealed that an expectancy bias is not restricted to gains in IQ, but may affect the learning of motor skills as well.

Further studies in which the researcher manipulated the expectancy bias, support the theory that teacher expectations affect pupil performance. Beez (1968), working with adult tutors who were teaching Headstart children, Schrank (1968), working with Air Force mathematics instructors teaching air-men mathematics, Meichenbaum, Bowers, and Ross (1969), working with the teachers of institutionalized adolescent female offenders, all reported similar results. In all of these cases, the teachers were led to believe that the experimental group either had a high learning potential or a low learning potential. Even though there was no factual basis for these expectations , in each case the students of the teachers with high expectations learned more than the students of the teachers who expected little.

An element which is common to all of the studies previously cited is that the expectations of the teachers were experimentally manipulated. Other studies have been conducted which focus on naturally occurring teacher expectations.

Palardy (1969) experimented with two groups of first grade teachers and their effects on the reading achievement of their pupils. One group of teachers believed that in the first grade boys were as successful as girls in learning how to read. The second group of

teachers believed that girls were more successful than boys in learning how to read. A total of ten teachers were selected for the sample. The ten classes used in the study were given a reading readiness test in September. The scores indicated no initial difference between the two groups. Upon retesting for reading achievement in March, however, differences were revealed among the boys. Those boys whose teachers believed they could read as well as girls averaged 96.5 on the tests, while those boys who weren't expected by their teachers to be as successful as girls averaged 89.2 on the tests. The girls averages were 96.2 and 96.7 respectively. The difference in boys' mean scores closely approximated significance at the $p \leqq .08$ level.

An investigation similar to that of Palardy's was that of Doyle, Hancock, and Kifer (1971). They related estimated IQs, reading achievement, and actual IQs of children in the first grade. It was found that teachers tended to overestimate the IQs of girls and underestimate the IQs of boys. Despite the fact that an IQ test showed that there was no significant difference between the boys and girls in the study, the girls scored higher on a reading achievement test. Furthermore, any overestimation of IQ, regardless of sex, related directly to a high score in reading achievement. Whether the teacher was greatly influenced by the reading achievement of pupils when estimating their IQ or whether her expectations for their reading performance affected her style of teaching reading was not known. The researchers, however, tend to support the latter reason as a explanation of their findings.

The research of Mackler (1969) and Douglas (1964) illustrate how expectations have become "institutionalized" in the educational system. Both men studied the tracking system used in a Harlem school and in British schools, respectively. Their findings show that: (1) "teachers' expectations about a child's achievement can be affected by factors having little or nothing to do with his ability," (i.e. race, color, sex, socio-economic status, etc.); "yet, (2) these expectations can determine his level of achievement by confining his learning opportunities to those available in his track." (Brophy and Good, 1973:70)

One further example of institutionalized expectations is the use of cumulative records. Based on his teaching experience in a ghetto school, Herbert Kohl came to the following conclusion: "It is amazing how 'emotional' problems can disappear, how the dullest child can be transformed into the keenest, and the brightest into the most ordinary when the prefabricated judgements of other teachers are forgotten". (1968:13)

2. PROPHECY COMMUNICATION

According to Brophy and Good (1973) the mere presence of an expectation alone is not, in itself, self-fulfilling. Rather, it is the behavior resulting from the expectation that leads it to be self-fulfilling. The following model suggests the process by which this phenomenon works in the classroom:

(1) The teacher expects specific behavior and achievement from particular students.

(2) Because of these different expectations, the teacher behaves differently toward the different students.

(3) This teacher treatment tells each student what behavior and achievement the teacher expects from him and affects his self-concept, achievement motivation, and level of aspiration.

(4) If this teacher treatment is consistent over time, and if the student does not actively resist or change it in some way, it will tend to shape his achievement and behavior. High-expectation students will be lead to achieve at high levels, while the achievement of low-expectation students will decline.

(5) With time, the student's achievement and behavior will conform more and more closely to that originally expected of him." (Brophy and Good, 1973:75)

Assuming this model to be valid, then a problem which remains to be explored is the relationship of teacher behavior to prophecy communication. In what ways does a teacher unintentionally communicate her expectations to her pupils? In an attempt to explain this, Rosenthal has proposed a "four-factor theory" (Rosenthal, 1973:60). The four factors are climate, feed-back, input, and output.

The climate factor refers to the socio-emotional mood which the teacher creates during interaction with her students. When this interaction is with "special potential" students, Rosenthal suggests that the teacher behaves in a more friendly, warm, and supportive way. This behavior is exhibited mainly in a nonverbal manner (smiles more, nods approval, maintains longer eye contact and closer body positioning).

Adair and Epstein (1967) were able to show that auditory cues alone can result in the communication of an expectation. In their study they used tape-recorded instructions which were given to the subjects during a perception test. Two groups of recordings were

made: one of a group of experimenters who expected success perceptions for their subjects and the other of a group of experimenters who expected failure perceptions for their subjects. The group who expected success perceptions from their subjects, evoked success and the group who expected failure perceptions evoked failure. Apparently experimenter expectations were communicated effectively through inflection and intonation. This influence of verbal cues in the covert communication of an expectation is also supported by Rosenthal and Fode (1963).

A study in a similar vein (Conn, Edwards, Rosenthal, and Crowne, 1968) was able to measure children's accuracy in judging the emotional tone of their teacher, an adult female. Those children who were more accurate benefited more from favorable teacher expectations.

Rosenthal and Fode (1963) and Cooper (1971) have found evidence which suggests that visual cues may also function in the transmission of an expectancy. In Cooper's study, the subjects were made to feel successful or unsuccessful depending upon the experimenter's use or avoidance of eye contact. This, in turn, affected the subject's performance on a neutral task.

Rosenthal's second factor, feedback, though closely related to climate, differs from it for it depends upon student responses and specifically refers to active teaching. Brophy and Good (1970) found that high-achieving students received more praise and support from their teachers and more feedback generally, whether or not they gave correct responses to questions. In their study, teachers ignored

fifteen percent of the answers from low-achievers but only three percent from high-achievers. Wright and Nuthall (1970) were also able to give support to the feedback factor for they found that saying 'good' and thanking a pupil were directly related to achievement. From Rosenthal's own review (1973) of the studies conducted on feedback, he implies with 'modest certainty' that teacher praise is one determinant of pupil achievement but the role of critism is vague.

The theory that teachers actually teach more to students for whom they hold high expectations is the inference made by Rosenthal's input factor. Although very little research has been done in this area, the experiment conducted by Beez (1968) gives evidence to support this theory. Beez worked with sixty teachers who were teaching symbol learning to sixty preschool children in a summer Headstart program. Half of the teachers were led to believe that their pupils would be good in symbol learning and the other half were led to believe that their pupils would be poor. What Beez discovered was that there were very overt differences in the teaching styles of the two groups of teachers. The teachers who had favorable expectations for their pupils taught a significantly greater number of symbols than did the teachers with unfavorable expectations. A post-test revealed that the children in the 'high-ability' group had a mean of 5.9 symbols learned while the low-ability' group had a mean of 3.1 symbols learned. Beez concluded that when teachers

> expect the child to do poorly they attempt to teach less, spend more time on each task, give more examples of meaning, and repeat the task more often than when they expect better performance from the child. (Beez, 1968:606)

Finally, the output factor is related to the number of opportunities which a student has to respond and to question. Rowe, and Brophy and Good have studied this factor and have found that teachers react to high-achieving pupils in ways which perpetuate their success. Teachers would wait longer for a response from a high-achieving student than they would from a low-achieving student (Rowe, 1969), and it would more often be the high-achievers who would be given a second chance to respond after an incorrect answer. Although there is insufficient evidence to support or refute Rosenthal's theory, he suggests that:

> teachers encourage greater responsiveness from students of whom they expect more. They call on such students more often, ask them harder questions, give them more time to answer and prompt them toward the correct answer. (Rosenthal, 1973:62)

3. TEACHER EXPECTANCY AND STUDENT SELF-CONCEPT

Although very few studies have focused on the effect of a teacher's expectations on the self-concept of the student, this phenomenon warrants some attention. A major contribution in this area stems from the work of Pitt (1956). He was unable to support the hypothesis that a teacher's expectations become self-fulfilling prophecies, however, he did find a relationship between the arbitrary increase or decrease of pupils' IQs and pupil self-ratings. When compared with those boys whose IQs had been ficticiously increased, the boys with lowered IQs felt that they didn't work as hard, that school was more difficult, that their teachers marked them harder, and that they didn't enjoy school as much. Mackler (1969) and Douglas (1964), who studied

ability grouping and tracking systems, support Pitt's findings for they felt that the self-concept and aspirations of any child who was needlessly placed in a low track, deteriorated over the years.

Brookover, one of the first sociologists to conduct research on the educational self-fulfilling prophecy, found that a student's academic self-rating predicted his performance, holding measured intelligence constant. (1962) The student's self-concept of ability may differ from subject to subject and is positively related to the image he perceives significant others (parents, teachers and peers) hold of him.

> It seems clear that self-concept of ability functions independent of measured intelligence in predicting school achievement. If self-concept is subject to modification, as theoretically postulated, and if modification in the images and expectations which other hold for a student takes place, the significant enhancement of achievement may be possible. (Brookover, 1962:75)

CHAPTER III

METHODS AND PROCEDURES

1. SELECTION OF SAMPLE

Since the major focus of this study was to determine the effectiveness of the methodology, only one swimming teacher and two of his classes were selected for the sample. The two classes consisted of 18 boys and girls, ages 5 - 10 years, who were receiving instruction at the Junior Red Cross level at the University of Alberta. Each class had a forty-five minute period of instruction, each afternoon, five days a week, for two weeks.

The teacher was a physical education student who was a qualified Red Cross Instructor employed by the aquatics department at the University of Alberta.

2. SELECTION AND CONSTRUCTION OF INSTRUMENTS

SWIMMING SKILLS TEST

The Swimming Skills Test was designed to measure the improvement in swimming skill of the students as a direct result of the instruction they received during the two weeks. Five items were included on the test and these were based on the skills listed on the Junior Red Cross test sheet. These items were: distance swim on the front, distance swim on the back, (stroke style was not a criterion for evaluation on these two items) a dive from the pool deck, treading water and drownproofing. Students were scored on each item on a scale ranging from one to five points, with each point designating a specific area of skill ability.

In a pilot study, this test was administered to eight students on the first day of instruction in a Junior Red Cross class. It was found that all of these students had been overtaught at the beginner level. As a result, they scored in the upper range on the skills test and therefore, a post-test using the same scoring system would not measure improvement. Based on these findings, a new method of scoring was devised. To account for skill improvement for all subjects, regardless of their swimming ability at the beginning of the course, a maximum performance test was administered as the pre-and-post-skills test. The scoring system was then constructed which would allow for the evaluation of the improvement in each skill item for each subject. (See Appendix A for scoring system).

STUDENTS'S SELF RATING OF SWIMMING ABILITY

The student's self-rating of swimming ability was a test designed to measure the student's own rating of his swimming ability. It was constructed as a paper-pencil test and the students were requested to rate themselves on each of five items: distance swim on the front, distance swim on the back, a dive from the pool deck, treading water and drownproofing. The students rated themselves as excellent, very good, good, fair or unsatisfactory. The keyed weights for these evaluations were as follows: excellent-5 points; very good-4 points; good-3 points; fair-2 points; unsatisfactory-1 point.

Students were requested to write this test from three different reference points. On the first form, the students rated themselves according to how well they felt they could swim at that point in time.

Secondly, they were asked to give a potential rating; i.e., how well they felt they would be able to swim at the end of their swimming course. Third, they were asked to rate themselves according to how they perceived their teachers would rate them at that point in time.

TEACHER RATING OF STUDENT'S SWIMMING ABILITY

The teacher rating of student's swimming ability was a test designed to measure the teacher's subjective rating of a student's swimming ability. It was exactly the same test which the students used to give a self-rating of their swimming ability; i.e., the teacher rated a student on each of the five skill items (distance swim on the front, distance swim on the back, a dive from the pool deck, treading water, and drownproofing) as being excellent, very good, good, fair, or unsatisfactory. The keyed weights for these ratings were as follows: excellent-5 points; very good-4 points; good-3 points; fair-2 points; unsatisfactory-1 point.

TEACHER OBSERVATION SCHEDULE (TOS)

For the purpose of observing the specific teacher behaviors of feedback and reward, correcting and prohibiting, a portion of the TOS (Rushall, 1973) was used. Although in its complete form the TOS classifies teacher behavior into seven categories, only the first two categories were required to fulfill the needs of this study. The descriptions of these two categories as defined by Rushall (1973) are as follows:

BEHAVIOR CATEGORY 1

1. Feedback (F): The teacher provides information in order to tell the pupil that his performance was satisfactory and that he can continue further. The nature of the information is such that it indicates either of two things, 1) the performance was satisfactory and should be repeated in the same manner, or 2) the performance was satisfactory but can be improved even further by incorporating additional features which are included in the feedback communication. Feedback can concern both skill and general behaviors and must stipulate what to do on the next occurrence of the behavior.

Examples: "That was a good arm extension. Try and keep it that way." " Your body position is almost perfect. Drop your head a little lower and it will be perfect." Your knowledge of the rules is very good. A little more experience at judging and you will be a very competent judge."

2. Rewarding (R): The teacher openly demonstrates pleasure with the behavior of a pupil, group or class. It conveys a positive feeling about or acceptance of the behavior to which it is related. It can be verbal (e.g. "Good!"; "That is the finest technique I have seen you do"; "Great effort") or non-verbal (e.g. excited clapping, flashing a victory sign). The reward is differentiated from feedback in two ways, 1)it is purely directed at some past performance, and 2) the information content does not indicate what to do on the next occurrence of the behavior.

BEHAVIOR CATEGORY 2

1. Correcting (C): The teacher provides information in order to tell the pupil that his performance was not satisfactory and how it must be altered to continue further. The content should include the performance characteristics which must be introduced to produce at least a satisfactory performance. This contrasts with feedback as correcting implies that the performer still has to achieve an adequate performance. Correcting can concern both skill and general behaviors and must stipulate what to do on the next occurrence of the behavior.

Examples: "Your racquet head was too low. Raise it up to a position higher than your wrist when you play the next backhand." "Your time was poor. Next time try to keep an even pace throughout the whole race rather than going out very fast."

2. Prohibiting (P): The teacher disciplines or openly displays displeasure with the behavior of a pupil, group or class. It conveys a negative feeling about or unacceptability of the behavior to which it is related. It can be verbal (e.g. "That was a poor shot") or non-verbal (e.g. "thumbs-down" sign, halt hand signal). Prohibiting is differentiated from correcting in two ways, 1) it is purely directed at some past performance, and 2) the information content does not indicate what to do on the next occurrence of the behavior.

The selection and administration of the test instruments was directly influenced by the attempt to approximate, as closely as possible, the type of methodology which had been used in classroom research. The purpose of this was to investigate the effectiveness of this methodology in a physical education setting.

3. COLLECTION OF THE DATA

The methods used for the collection of the data consisted of a practical swimming skills test, paper-pencil tests designed to give subjective ratings of swimming ability and the observation of specific teacher behaviors.

After the sample was selected, the first step was to attempt the manipulation of the teacher's expectations. The teacher was informed that a series of tests had been devised which would distinguish those students who had the greatest potential for improvement in his swimming classes. These tests were to be administered during the first session of each of the classes being taught by the teacher. Based on the results of these tests, one class was to be designated as the "special potential" group for it was the one which would show the greatest overall improvement in swimming over the two-week instructional period.

The teacher was led to believe that the purpose of this was to compare the behaviors of the "special potential" group and the control group in an attempt to investigate the relationship between student behavior and skill improvement. He was informed that he would not be told which specific student behaviors were being observed because, in this way, he would not be able to influence the frequencies of these behaviors.

The pre-testing was conducted during the first session of each class. The swimming skills test was administered to each student and the performances were recorded by a panel of experts. The experts, two active Red Cross instructors were asked to record the maximum performance of each student on each of the five skill items. The subjects were instructed to do their "best" on each of the skill items. Because the self-rating tests were to be administered immediately following the skills test, the examiners were informed that they were not to give the students any feedback, verbal or non-verbal, related to their performance.

Immediately following the practical test, the self-rating of swimming ability tests were administered to the students. First, they were asked to rate themselves according to how well they felt they were able to swim at that point in time. These tests were collected and the second set were given to the students. On the second set the students were asked to give the rating which they perceived their teacher would give them. These tests were collected and on the final set the students were asked to rate their potential

swimming ability, i.e. how well they felt they would be able to swim at the end of the course.

The teacher was not present during these written tests and was given no information related to the content of the tests, for he was led to believe that these were the psychological tests which would distinguish the "special potential" group.

Following this testing session, one of the two classes was randomly selected as the "special potential" group. Before the next class, this information was told to the teacher as a "matter of interest". He was not aware, however, that the selection of the "special potential" group had been random.

Throughout the two weeks of the course, the researcher was in attendance at every lesson. It was felt that this would reduce any influence the researcher may have on the behavior of the teacher or the students. Fifty percent of these classes, in both groups, were chosen for observation of teacher behavior. These were the fourth to eight classes inclusive. The observational periods were fifteen minutes in length, beginning five minutes after the class had started. These first five minutes were allowed for late-comers to join the class and/or for administrative duties which may be carried out by the teacher. During the observations, the researcher tallied all occurrences of the specific teacher behaviors of feedback, reward, correcting and prohibiting.

In order to insure the researcher's reliability and objectivity another person was trained as an observer at the same time as the

researcher. The training sessions, conducted previous to the actual testing, consisted of learning the four categories and discussing which teacher behaviors would be included in these categories. This was followed by an observation and discussion of a film segment of a teacher instructing a swimming class. When both observers felt competently trained, simultaneous and independent observations of another film segment were made. The percentage of agreement between the two observers was calculated using the following formula:(Rushall, 1973)

$$\frac{NA1 + NA2 \ldots\ldots + NA_n}{NA1 + ND1 + NA2 + ND2 \ldots\ldots + NA_n + ND_n} \times 100 = \% \text{ agreement}$$

Where NA_n = number of agreements for category n, and

ND_n = number of disagreements for category n

The usual acceptance level for inter-observer percentages of agreement is 80%. (Rushall, 1973:3)

For this first session, just prior to testing, the observers agreed 85% of the time time.

During the actual testing, the third session of one of the groups to be used for observation was filmed on a videorecorder. Subsequent analysis of this film segment revealed a 90% agreement between the researcher and the other trained observer.

The final stages of testing took place during the tenth lesson. At this time, the skills test was again administered to the students and their performances were recorded by the panel of experts. Follow-

ing the skills test, the students were given two of the three self-rating tests which had been used in the pre-testing session. These two were from the reference points of how the students would rate their swimming ability at that point in time and, how they felt their teacher would rate them.

The next stage of data collection took place immediately following the tenth and final lesson. At this time the teacher was asked to give his rating of the swimming ability of each of his students in his two classes.

One week after the completion of testing, a debriefing session was held with the teacher. This was to determine how successful the researcher's manipulation of the teacher's expectations had been and to assess the teacher's knowledge of the purpose of the study and to get his general impressions of what had actually taken place.

4. ANALYSIS OF DATA

The statistics used to analyze the data were the student t-test and the Product-moment correlation coefficient. The t-test was used to determine significant differences between the two groups and the Product-moment correlation was used to determine the degree of the relationship between the variables. The correlation coefficients were transformed to z_r scores to determine the significance of the differences between the correlations.

CHAPTER IV

RESULTS AND DISCUSSION

The results have been divided into four main sections. In the first three sections, the results of the three sub-problems have been presented and followed by a discussion of these results. The fourth section is devoted to a discussion of the methodology used in the study and its possible effects on the results.

1. EXPERIMENTAL MANIPULATION OF TEACHER EXPECTATIONS

During the week previous to the beginning of testing, the teacher to be used in this study was approached. At this time, permission to use the teacher's two Junior Red Cross classes in the experiment was requested and granted. The teacher was told that a set of psychological and swimming skill tests had been devised which would determine which one of his two classes would show the most skill improvement over the two weeks of the course. These tests were administered to each class on the first day of the course. The teacher had no knowledge of the content of the purported psychological tests, but he was present during the skills test.

Before the second day, one class was randomly selected by the researcher and designated as the "special potential" group. The teacher was given this information, however, he was not aware that the selection had been random and was led to believe that it was based on fact. His response when he found out which class had been designate as the "special potential" one was "Yah, you're right." Because of this re-

sponse, which was entirely unexpected at this point, no further mention was made by the researcher throughout the testing sessions of which class was the "special potential" group. It was assumed that a teacher bias had formed already without the attempts by the researcher to manipulate the teacher's expectations.

It is important to mention that the teacher knew the researcher fairly well and therefore, this diminished any distrust on the part of the teacher for what the researcher was purporting to be doing.

Throughout the two weeks of instruction the researcher was in attendance at every class. The teacher was informed that during five of these classes, the researcher would be observing and recording frequencies of specific student behaviors (when actually, specific teacher behaviors were being observed). The teacher was given no information as to what 'student' behaviors were being observed. The reason for this was so that "he would not be able to unconsciously influence the frequencies of these behaviors". He was satisfied with this reasoning. One of these observation session was filmed on a videorecorder for subsequent analysis with another trained observer.

At no time during the testing did the teacher appear to be uncomfortable or threatened by the presence of the researcher. This proved to be valuable for it allowed the researcher to make general observations of the teacher's genuine interaction with each group.

It became more and more obvious as the course progressed, that the teacher experienced the children of the control group as being far more enjoyable and satisfying to work with. As a group, they appeared to be more attentive and cooperative, whereas the "special potential" group had two children who proved to be distracting

influences and thus tended to disrupt the class quite frequently. It was not uncommon, especially during the second week of instruction, for the teacher to express to the researcher the pleasure he experienced when teaching the children of the control group.

On the last day of the course, each class was again given a series of tests which were used during pre-testing with the exception of the student's self-rating of potential swimming ability. Immediately following the administration of these tests, the teacher was asked to rate the swimming ability of each of his students in both classes.

One week following the completion of testing a debriefing session was held with the teacher. He was first asked to give to the researcher all the knowledge he had related to the purpose of the study. To his knowledge, the researcher was investigating the relationship of some student behaviors to improvement in swimming. The pre-post skills test had been used to test for improvement and the observation sessions had been used to record frequencies of certain student behaviors, of which he was unaware. He further stated that the reason for the filming of one session was so that it could be analyzed by the researcher and another observer. He expressed feelings of discomfort during the filming and felt that this had changed his teaching behavior somewhat, i.e. he was very aware of 'trying' to be good when teaching. This was all the information he gave, even after attempts at prompting him to determine if he could remember anything else.

When reminded that he had been told which one of his two classes would show the most improvement, he said that he could not remember being told this and furthermore, he did not remember agreeing with the researcher at the time.

When told, at this point, which one of his classes had been designated as "special" his response was: "That class?" Also, when told that it had not been student behavior that was being observed but, rather, teacher behavior, he reacted with disbelief. At no time was he ever aware that he was being observed.

Further debriefing included the results of the study and a discussion of them.

DISCUSSION

The main purpose of this case study was to determine the feasibility of a researcher biasing a teacher's expectations for his students, expecially in the physical education setting. In this matter, the results of this study are not in agreement with Rosenthal, who has contributed the major portion of research in this area. It would appear that in his Oak School experiment (1968) and others similar to this, Rosenthal was able to achieve some degree of success in manipulating teacher expectations. An explanation for the difference between this present study and Rosenthal's can be related to the information about the student upon which the expectations were formed. Rosenthal gave the teachers involved in his study fictious student IQs and then studied the gains made in IQs. Information related to IQ is difficult for the teacher to dispute for the IQ is not readily observable and thus contrived values are difficult to dispute. A study such as the present one, however, bases the manipulation of teacher expectations on information related to the skill ability

(i.e. swimming) of the students. Any fictious information given in this type of situation is more easily disputed for the teacher is able to actually observe and determine to a great extent the abilities and improvements of the students. Thus, an attempt to bias teacher expectations may be met with some resistance if they are incongruent with the teacher's own observations.

In the one other study which has been conducted in a physical education environment, Burnham was able to successfully manipulate the expectations of swimming teachers at a summer camp. He did this, however,by using two groups of teachers working with two groups of swimming classes, one of which was designated as the "special potential" group. As a result, the teachers were exposed only to the children of the "special potential" group or the control group, depending upon which group had been assigned to them. For this reason there was no opportunity for comparing the swimming abilities of the students in the different groups and thus there was no reason to dispute the biases for their students which had been given to them by the researcher. Unfortunately, the results of the study cannot be conclusively shown to be the effects of the teacher expectations for the individual differences, abilities and experience of the teachers involved may also account for the differences in the two groups at the end of the swimming course.

Thus, the apparent ease with which Rosenthal was able to bias teacher expectations is not apparent in those studies in physical education settings. The important variable which must be considered

here is the personality of the teacher. His teaching experience, self-confidence and degree of gullibility determine to some extent his acceptance of a researcher's bias. Also the basis of the information about the student and from whom this information is received are important factors controlling the successful manipulation of teacher expectations.

2. EFFECTS OF TEACHER EXPECTATIONS ON TEACHER BEHAVIOR

Table 1. MEAN DIFFERENCES BETWEEN THE TEACHER RATINGS AND THE EXPERT RATINGS

Rater	Control	Special Potential	Mean Difference
Teacher	15.22	12.89	-2.33
Expert	14.44	11.11	-3.33
Mean Difference	.38	1.78	

one-tailed test; $p \geq .05$

Hypothesis A. The teacher's ratings of swimming ability will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the teacher would rate the "special potential" group significantly higher than the control group. Rather, the teacher's ratings favored the control group but the difference is not significant.

Hypothesis B. The difference between the teacher's ratings and the expert's post ratings will be significantly greater for the "special potential" group than for the control group.

Although the difference between the teacher's ratings and the expert's ratings is in the predicted direction, this difference is not significant and therefore the hypothesis is rejected.

Table 2. FREQUENCIES OF SPECIFIC TEACHER BEHAVIORS

	Feedback	Reward	Total	Correcting	Prohibiting	Total
Control	11	71	82	144	12	156
Special	12	66	78	135	9	144
Difference	1	-5	-4	9	3	12

Hypothesis C. The frequencies of feedback and reward behaviors exhibited by the teacher will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the teacher would give significantly more feedback and reward to the "special potential" group. Rather, the frequencies favor the control group but the tendency is not significant, therefore, the hypothesis is rejected.

Hypothesis D. The frequencies of correcting and prohibiting behaviors exhibited by the teacher will be significantly greater for the control group than for the "special potential" group.

Although the frequencies of correcting and prohibiting occur in the predicted direction, the difference between the frequencies for the two groups is not significant and therefore, the hypothesis is rejected.

Table 3. PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN TEACHER RATINGS AND OTHER RATINGS OF STUDENT SWIMMING ABILITY

Other Ratings	Teacher Ratings		Difference (Z_r)
	Control	Special Potential	
Expert	.46*	.63**	.27
Self-rating (Teacher)	.38	.17	.39
Self-rating (Post)	.24	.53*	.65*

Z_r - Transformed r scores
One-tailed test; df=18
* = $p \geq .05$
** = $p \geq .005$

Hypothesis E. The correlation between the expert rating and the teacher rating of swimming ability will be significantly greater for the control group than for the "special potential" group.

The data fails to support the hypothesis that the greater percentage of agreement between the teacher and the expert would occur in their ratings of the control group. Instead, the data reveals that they agreed more often in their ratings of the "special potential" group but this difference is not significant and thus the hypothesis is rejected.

Hypothesis F. The correlation between the student self-ratings, from the reference point of the teacher, and the teacher ratings will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the greater percentage of agreement between the student perceived teacher ratings and the teacher ratings would occur in the "special potential" group. Although neither correlation is significant, the higher percentage of agreement occurred in the control group and therefore, the hypothesis is rejected.

Hypothesis G. There will be no significant difference in the correlation between the post self-ratings and the teacher ratings for the control group and the "special potential" group.

The data fails to support the hypothesis, for a significant difference was revealed between the correlations for the control group and the "special potential" group . Since the "special potential" group had a significantly higher correlation between post self-ratings and teacher ratings, the hypothesis is rejected.

DISCUSSION AND INTERPRETATION

The purpose of this section was to determine if the teacher behaved differentially toward the "special potential" and the control groups. When the teacher ratings of the students' swimming ability were compared with the non-biased expert ratings, it was found that there was no significant difference between the two. Furthermore, the two ratings showed significant positive correlation for both groups. Evidently, the teacher was not biased in his ratings of the control group or the "special potential" group.

To check out further the possibility of the existence of a teacher bias, the frequencies of feedback, reward, correcting and prohibiting behaviors exhibited to each group were compared. The underlying assumption here, which stems from the review of studies conducted in this area, was that the teacher would give more feedback and reward but less correcting and prohibiting to the "special potential" group than would be given to the control group.

Again, the teacher did not interact more favorably with one group as was expected. Therefore, if any teacher bias, manipulated or

not, existed, it was not revealed through the observations of the frequencies of feedback, reward, correcting and prohibiting, for essentially the two classes were treated equally in this respect.

Assuming that teacher treatment of a student will affect that student's self-rating, an interesting finding was revealed when the correlations between post self-ratings and teacher ratings for the control group and the "special potential" were examined. It had been predicted that there would be no difference between the two correlations. This was based on the fact that the control group and the "special potential" group had never been exposed to each other, therefore, they were not given the opportunity to compare differential teacher treatment, if, in fact, it existed. As a result, neither group could dispute the communication between the teacher and themselves. If the teacher expected that a group was 'average' or 'special' and treated the group as 'average' or 'special', chances are the group would think of themselves as being 'average' or 'special' and therefore, the teacher ratings and the self-ratings should be highly correlated.

This reasoning, however, was not supported by the data. Due to the fact that there was a significant difference between the correlations for the two groups, in favor of the "special potential" group, it appears that the teacher was more effective in communicating his expectations to the "special potential" group.

An explanation of this may be related to the fact that the teacher found the control group a far more enjoyable group to teach. His frequent displays of pleasure with the group's performance may have been

interpreted unrealistically by the students which resulted in a greater discrepancy between the students' self-ratings and the teacher ratings than was evident in the "special potential" group.

3. EFFECTS OF TEACHER EXPECTATIONS ON STUDENT SELF-RATING AND SKILL IMPROVEMENT

Table 4. PRODUCT-MOMENT CORRELATIONS BETWEEN THE SELF-RATINGS AND THE EXPERT RATINGS OF STUDENT SWIMMING ABILITY

Ratings Compared		Control	"Special Potential"	Difference
Self-rating and Expert rating	Pre	.41*	.14	.51***
	Post	.32	.03	.57***
Self-rating (Teacher) and Self-rating	Pre	.88****	.69****	.91****
	Post	.67****	.90****	1.41****
Self-rating (Potential) and Post Self-rating		.29	.62**	.74****

One-tailed test; df=18
*=p≥.05
**= p≥.025
***=p≥.01
***=p≥.005

Hypothesis H. The correlation between the self-ratings and the expert ratings will be significantly greater for the control group than for the "special potential" group.

The results indicated that in both the pre-test and post-test situations the hypothesis was supported. The self-ratings and the expert ratings showed a significantly higher percentage of agreement in the control group than in the "special potential" group.

Hypothesis I. The correlation between the self-ratings, from the reference point of the teacher, and the self-ratings of swimming ability will be significantly greater for the "special potential" group than for the control group.

The results indicated that in the post-test situation the "special potential" group had a significantly higher percentage of agreement between the perceived teacher ratings and the self-ratings than the control group. Therefore, the hypothesis was accepted.

Hypothesis J. The correlation between the potential self-ratings and the post self-ratings will be significantly greater for the "special potential" group than for the control group.

The results indicated that the hypothesis was supported. The "special potential" group showed a significantly higher percentage of

Table 5. COMPARISON OF THE MEAN DIFFERENCES IN SELF-RATINGS AND THE EXPERT RATINGS

Ratings		Control	Special Potential	Difference
Self-rating	Pre	15.11	15.33	.22
	Post	18.	18.22	.22
Self-rating (Teacher)	Pre	16.44	15.	1.44
	Post	18.22	17.89	.33
Self-rating (Potential)		20.33	18.55	1.78
Expert Rating	Pre	9.56	9.22	.34
	Post	14.44	11.11	3.33

One-tailed test
$p \geq .05$

agreement between the self-ratings of potential swimming ability and the post self-ratings than did the control group.

Hypothesis K. The post self-rating of swimming ability will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the "special potential" group would rate their swimming ability higher than the control group at the end of the course. The mean difference is not significant and therefore, the hypothesis is rejected.

Hypothesis L. The post self-ratings, from the reference point of the teacher, will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the perceived teacher ratings of the "special potential" group would be significantly greater than those of the students in the control group. The mean difference is not significant and therefore, the hypothesis is rejected.

Hypothesis M. The difference between the potential self-ratings and the post self-ratings will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the "special potential" group would show a significantly greater difference between the potential self-ratings and the post self-ratings than the control group. The mean difference is not significant and therefore, the hypothesis is rejected.

Hypothesis N. The improvement in swimming ability will be significantly greater for the "special potential" group than for the control group.

The data fails to support the hypothesis that the "special potential" group would show significantly greater improvement in swim-

ing ability than the control group. The mean difference is not significant and therefore, the hypothesis is rejected.

DISCUSSION AND INTERPRETATION

The purpose of this section was to examine what effects, if any, the teacher's expectations had on the students' self-ratings of and improvement in swimming ability.

An examination of the correlation between self-ratings and expert ratings (although there was no significant difference in these ratings for each group) revealed a significant difference between the control group and the "special potential" group, with the control group having the higher percentage of agreement. It appears then, that the control group had a more realistic outlook of their actual swimming ability than did the "special potential" group, although the correlation between self-rating and expert rating is not significantly high for the control group.

The results further indicate that the correlation between students' perceived teacher ratings and the self-ratings (despite the fact that there was no significant difference between these ratings when the groups were compared) was significantly greater for the "special potential" group than for the control group at the end of the course. Thus, it can be inferred that the students of the "special potential" group were more confident that their teacher would rate them the same as they would rate themselves.

Furthermore, although there was no significant difference between the groups on the potential self-ratings or the post self-ratings, the correlation between these two ratings was significantly greater for the

"special potential" group. Thus, it can be inferred that the "special potential" group felt that they had reached the potential level of swimming ability which they had predicted at the beginning of the course.

A review of the results of this section reveal an inconsistency. On one hand, all of the ratings of the students' swimming ability showed no significant differences between the two groups. On the other hand, however, the correlations using combinations of different ratings, indicated that there were significant differences favoring the "special potential" group, especially when self-ratings and teacher ratings were correlated. What this evidence seems to indicate was that the self-fulfilling hypothesis was actually having an effect, although not to the extent that had been predicted. The teacher unconsciously transmitted to the students of the "special potential" group expectations that were higher than those transmitted to the students of the control group.

4. DISCUSSION OF THE RESULTS

The results of this study indicate that the expectations of the teacher had not been successfully manipulated by the bias given to him by the researcher. This failure to induce the desired expectations should not necessarily be taken as rejection of the self-fulfilling prophecy, although the negative results revealed in the analysis of the data would tend to support this proposition. Rather, the negative results are more likely due to the researcher's failure to manipulate the expectations of the teacher than to the failure of teacher expectations to affect teacher behavior and, subsequently, student behavior.

Analysis of the data revealed that the comparisons of the mean differences between the expert ratings, self-ratings and teacher ratings, and the frequencies of feedback, reward, correcting and prohibiting for the control group and the "special potential" group indicate that there were no significant differences between the two groups. However, when the correlations between the expert ratings, self-ratings and teacher ratings for the two groups were compared, in most cases, very significant ($p \leq .01$) differences between the control group and the "special potential" group were revealed. These differences seem to indicate that the self-fulfilling prophecy was actually working and that the teacher unconsciously transmitted differential expectations to the two groups which favored the "special potential" group.

5. METHODOLOGICAL IMPLICATIONS

Since the major concern of this study was with methodology, this discussion will focus on the problems encountered during the testing and how the methodology may have affected the results.

One of the first problems to be dealt with was the attempt to manipulate the teacher's expectations without raising his suspicions regarding the testing procedures. It was felt, since the researcher and teacher were well-known to each other, that the information designating a "special potential" group did not have much of an impact on the teacher and as a result, did not affect his behavior. If the information had come from a higher authority figure, the manipulation of the teacher's expectations may have been more successful.

The fact that the teacher did trust the researcher proved to be valuable, however, during the observation of teacher behavior for at no time was the teacher suspicious of being observed. He was convinced that it was student behavior that was being observed.

Another reason which may account for the difficulty encountered in manipulating the teacher's expectations was the information upon which the manipulation was based. The decision to be made here was whether or not a whole class should be designated as "special potential" or only a few students within a class. Although both types of samples have been used in classroom research, the decision to use only a few students within a class predominates these studies. The reason for designating a whole class as "special potential" in this study was to avoid a further methodological problem, which would have been encountered during the observation of teacher behavior.

Since the object of observing teacher behavior was to compare specific behaviors exchibited to the control group and to the "special potential" group, it was necessary for each of these groups to be easily distinguishable, not only to the researcher but to the second trained observer as well. If both of these groups were in the same class, then the very nature of the swimming pool environment would necessitate marking the students in some way so that they could be easily distinguished . This was seen to be a disadvantage for it may introduce further variables which could confound the effects of the experimental treatment.

Based on these anticipated problems, it was decided that one class would be designated as "special potential" and the other as control. This was done at the risk of diminishing the credibility of the information given to the teacher. Taking this risk resulted in undesirable effects and thus this is one problem area which must be solved in future research.

Another factor which posed a problem was the time element. The students received instruction over a two week period only. The results showed that there was no significant improvement in swimming ability for either group. This would indicate that for this particular level of swimming ability, i.e. Junior Red Cross, that an extended period of instruction would be required to result in significant improvement. The important consideration here is the ability of the students at the beginning of the course. Burnham (1968) in his study of swimmers at a summer camp, revealed that non-swimmers showed significant improvement after only two weeks of instruction. Thus the length of the period of instruction and the swimming ability of the students are two related variables which should be controlled in further research.

CHAPTER V

SUMMARY AND CONCLUSIONS

1. SUMMARY

The central purpose of this study was to explore the feasibility of adapting a methodology used to investigate the "Pygmalion" effect in a classroom setting to a physical education setting, specifically, swimming instruction. A case study was conducted in which three main areas were investigated: 1) experimental manipulation of teacher expectations, 2) the effects of teacher expectations on teacher behavior, and 3) the effects of teacher expectations on the student's self-rating of and improvement in swimming.

Eighteen children, ages 5 to 10 years, and their teacher were used as subjects. The children formed two classes who were being given swimming instruction at the Junior Red Cross level at the University of Alberta. The teacher, a qualified Red Cross Instructor, was responsible for the instruction in both of these classes. Based on the results of purported psychological tests which supposedly determined swimming potential, the two classes were designated as the control group (n=9) and the "special potential" group (n=9). The teacher was led to believe that the "special potential" group was that class of students who possessed the greatest potential for improvement in swimming and that this class, when compared to the control group, would show significantly greater improvement.

The instruments which were used consisted of tests designed to obtain subjective ratings of a student's swimming ability, a practical swimming skills test, a teacher observation schedule to record frequencies

of specific teacher behaviors, and an informal discussion to debrief the teacher. The variables examined were: 1) the manipulation of teacher expectations, 2) teacher's use of feedback, reward, correcting and prohibiting behaviors, 3) student's self-rating of swimming ability, 4) student's self-rating of swimming ability, from the reference point of the teacher, 5) student's self-rating of potential swimming ability, 6) teacher's rating of student's swimming ability, and 7) expert's rating of student's swimming ability.

Based on the problems encountered during the testing and on the results obtained, the methodology was examined to determine its effectiveness in investigating the "Pygmalion" effect in a physical education environment.

2. CONCLUSIONS

Based on the results of this study, the following conclusions were made:

1. The methodology used to study the "Pygmalion" effect in a physical education setting was not effective. Methological changes may be needed to account for the nature of the environment and its variables.

2. Because the manipulation of the teacher's expectations was not successful, there is no conclusive evidence to support or refute whether teacher expectations effect teacher behavior and subsequently student's self-rating and improvement in swimming ability.

3. RECOMMENDATIONS

Due to the methodological difficulties encountered during this study, the following recommendations seem warranted:

(1) More effective techniques are required to insure that differential teacher expectations are, in fact, present.

Suggestions:

a. If two or more classes are to be used as subjects, the students should be streamed into either the control or special group after the pre-testing and on a basis that appears sensible to the teachers involved.

b. If both control and special potential subjects are in one class and, if differential teacher behavior is being observed, a method of distinguishing the control subjects from the special subjects is required. Whatever method may be selected, it is important to support it with a good rationalization which does not arouse teacher or student suspicion of the nature of the study.

c. Rather than manipulating teacher expectations, capitalize on natural teacher bias. This requires that the teacher have prior knowledge of the subjects and that he would choose the "special potential" group. Although this method has its advantages, it may be undesirable for the teacher expectations may be based on factual information rather than prejudices.

d. Manipulation of teacher expectations may be more effective if a test is designed which would give a "motor quotient" score, comparable to the intelligence quotient score. Each subject would then be assigned a score and a list of subjects and their scores would be given to the teacher.

This may serve as a constant reminder of who has the highest scores, i.e. the special students.

(2) The length and/or intensity of teacher exposure to the students should be greater than a two week period to allow for substantial teacher student interaction.

(3) If teacher behavior is a variable to be considered, a more extensive and sensitive instrument should be used while observing teacher behavior, in an attempt to determine how a teacher communicates his expectations to the students.

(4) Further research is required in a variety of physical education environments and with a greater number of teachers if the implications of the "Pygmalion"effect are to be understood.

BIBLIOGRAPHY

Adair, J. and J. Epstein, "Verbal Cues in the Mediation of Experimenter Bias", Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May, 1967. (Original not available) cited in W.L. Claiborn, "An Investigation of the Relationship Between Teacher Expectancy, Teacher Behavior and Pupil Performance", Unpublished doctoral dissertation, Syracuse University, 1968.

Amidon, E.J. and Ned A. Flanders. The Role of the Teacher in the Classroom: A Manual for Understanding and Improving Teachers' Classroom Behavior. Minnesota: Paul S. Amidon and Associates, Inc., 1963.

Anderson, D.F. and R. Rosenthal. "Some Effects of Interpersonal Expectancy and Social Interation on Institutionalized Retarded Children", Proceedings of the 76th Annual Convention of the American Psychological Association, 3: 479-480, 1968.

Anderson, G.J. The Assessment of Learning Environments: A Manual for the Learning Environment Inventory. Halifax: Atlantic Institute of Education, 1973.

Anderson, E. & J. M. Carlsmith. "Performance Expectancy as a Determinant of Actual Performance", Journal of Abnormal and Social Psychology, 65: 178-182, 1962.

Barrett, K.R. "A Procedure for Systematically Describing Teacher - Student Behavior in Primary Physical Education Lessons Implementing the Concept of Movement Education", Unpublished doctoral dissertation, University of Wisconsin, 1969.

Beez, W.V. "Influence of Biased Psychological Reports on Teacher Behavior and Pupil Performance", Proceedings of the 76th Annual Convention of the American Psychological Association, 3: 605-606, 1968.

Biddle, B.J. "Methods and Concepts in Classroom Research", Review of Educational Research, 37: 337-357, 1967.

Biddle, B.J. and W.J. Ellena (eds.). Contempory Research on Teacher Effectiveness. New York: Holt, Rinehart and Winston, 1964.

Bookout, E.C. "Teaching Behavior in Relation to the Socio-emotional Climate of Physical Education Classes", Research Quarterly, 38: 336-347, 1967.

Brookover, W.B. and David Gottlieb. A Sociology of Education. New York: American Book Co., 1964.

Brookover, W., A. Patterson and S. Thomas. "Self-concept of Ability and School Achievements", Office of Education, Cooperative Research Project No. 845, Bureau of Educational Research Services, College of Education, Michigan State University, 1962.

Brophy, J. and T.L. Good. "Teachers' Communication of Differential Expectations for Children's Classroom Performance": Some Behavioral Data", Journal of Educational Psychology, 61: 365-374, 1970.

Browne, P. "Verbal Interaction in Primary Reading Groups", Unpublished doctoral dissertation, University of Alberta, 1971.

Burnham, J.R. "Effects of Experimenter's Expectancies on Children's Ability to Learn to Swim", Unpublished master's thesis, Purdue University, 1968.

Claiborn, W.L. "An Investigation of the Relationship Between Teacher Expectancy, Teacher Behavior and Pupil Performance, Unpublished doctoral dissertation, Syracuse University, 1968.

Claiborn, W.L. "Expectancy Effects in the Classroom: A Failure to Replicate", Journal of Educational Psychology, 60: 377-383, 1969.

Clark, K.B. "Educational Stimulation of Racially Disadvantaged Children" In A.H. Passow (ed.) Education in Depressed Areas. New York: Teachers College, Columbia University, 1963.

Conn, L.K., C.N. Edwards, R. Rosenthal, and D.P. Crowne. "Emotion Perception and Response to Teacher Expectancy in Elementary School Children", Psychological Reports, 22: 27-34, 1968.

Cooper, Joel. "Self-fulfilling Prophecy in the Classroom: An Attempt to Discover the Processes by Which Expectations are Communicated", Final Report. New Jersey: Princeton University, 1971.

Dougherty, N. "A Plan for the Analysis of Teacher - Pupil Interaction in Physical Education Classes", Quest, Monograph XV, 1971.

Douglas, J. The Home and the School: A Study of Ability and Attaiment in the Primary School. London: MacGibbon and Kee, 1964.

Doyle, W., G. Hancock and E. Kifer. "Teachers' Perceptions: Do They Make a Difference?", Paper presented at the annual meeting of the American Educational Research Association, 1971.

Edwards, T.B. and A.B. Wilson. "A Study of Some Social and Psychological Factors Influencing Educational Achievement", (Mimeographed). Berkeley: University of California, Department of Education, June, 1961.

Engle, Mary. "The Stability of the Self-Concept in Adolescence", Journal of Abnormal and Social Psychology, 58: 211-215, 1959.

Engle, M. and W.J. Raine. "A Method for the Measurement of the Self-Concept of Children in the Third Grade", Journal of Genetic Psychology, 102: 125-137, 1963.

Ferguson, G.A. Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Co., Inc., 1971.

Fleming, E.S. and R.G. Anttonen. "Teacher Expectancy or My Fair Lady", American Educational Research Journal, 8: 241-252, 1971.

Flowers, C. "Effects of an Arbitrary Accellerated Group Placement on the Tested Academic Achievement of Educationally Disadvantaged Students", Unpublished doctoral dissertation, Columbia University Teachers College, 1966.

Galloway, C.M. "Nonverbal: the Language of Sensitivity", Theory into Practice, 10: 227-229, 1971.

Good, T.L. and J.E. Brophy. Looking in Classrooms. New York: Harper and Row, 1973.

Grant, B. and D. Hennings. The Teacher Moves. New York: Columbia University, Teachers College Press, 1971.

Gumpert, P. and C. Gumpert. "The Teacher as Pygmalion: Comments on the Psychology of Expectation", Urban Review, 3: 21-25, Sept., 1968.

Haas, H.I. and M.L. Maehr. "Two Experiments on the Concept of Self and the Reaction of Others", Journal of Personality and Social Psychology, 1: 100-105, 1965.

Hardyck, C.D. and L.F. Petrinovitch. Introduction to Statistics for the Behavioral Sciences. Philadelphia: W.B. Saunders Co., 1969.

Heiserman, M.S. "The Relationship Between Teacher Expectations and Pupil Occupational Aspirations", Unpublished master's thesis, Iowa State University, 1967.

Hoffman, D.E. and E.C. Cohen. "An Exploratory Study to Determine the Effects of Generalized Academic Performance Expectations Upon the Activity and Influence of Students Engaged in a Group Simulation Game", A paper presented at the American Educational Research Association Annual Meeting, Chicago, 1972.

Holt, J. How Children Fail. New York: Dell Publishing Co. Inc., 1964.

Kohl, Herbert. 36 Children. New York: New American Library, Inc., 1968.

LaBenne, W.D. and Bert I. Greene. Educational Implications of the Self-Concept Theory. California: Goodyear Publishing Co., Inc., 1969.

Lyon, H.C. Learning to Feel - Feeling to Learn. Ohio: Charles E. Merrill Publishing Co., 1971.

Mackler, B. "Grouping in the Ghetto", Education and Urban Society, 2:80-95, 1969.

Meichenbawn, D.H., K.S. Bower, and R.R. Ross. "A Behavioral Analysis of Teacher Expectancy Effect", Journal of Personality and Social Psychology, 13: 306-316, 1969.

Orlick, T. "An Analysis of Early Sports Participation", Unpublished doctoral dissertation, University of Alberta, 1972.

Palardy, J. "What Teachers Believe - What Children Achieve", Elementary School Journal, 69: 370-374, 1969.

Payne, D.A. "The Concurrent and Predictive Validity of an Objective Measure of Academic Self-Concept", Educational and Psychological Measurement, 22: 773-780, 1962.

Perkins, H.V. "A Procedure for Assessing the Classroom Behavior of Students and Teachers", American Educational Research Journal, 1: 249-260, 1964.

Perkins, H.V. "Factors Influencing Change in Children's Self-Concepts", Child Development, 29: 221-230, 1958.

Pitt, C.C. V. "An Experimental Study of the Effects of Teachers' Knowledge or Incorrect Knowledge of Pupil IQs on Teachers' Attitudes and Practices and Pupils' Attitudes and Achievement". Unpublished doctoral dissertation, Columbia University, 1956.

Robbins, S. "The Development of an Instrument to Analyze Teacher-Pupil Interaction in Teaching Elementary School Physical Education", Unpublished doctoral dissertation, University of Alberta, 1973.

Roethlisberger, F.J. and W.J. Dickson. Management and the Worker. Cambridge, Mass: Harvard University Press, 1939.

Rosenshine, B. "Evaluation of Classroom Instruction", Review of Educational Research, 40: 279-300, 1970.

Rogers, C.R. Freedom to Learn. Ohio: Charles E. Merrill Publishing Co., 1969.

Rosenthal, Robert. Experimental Effects in Behavioral Research. New York: Appleton, 1966.

Rosenthal, Robert. "Teacher Expectations", In G.S. Lesser (ed.) Psychology and the Educational Process. Glenview, Ill.: Scott, Foresman and Co., 1971.

Rosenthal, Robert. "The Pygmalion Effect Lives", Psychology Today, 7: 56-63, 1973.

Rosenthal, R. and J. Evans. Unpublished data, Harvard University, 1968. Cited in R. Rosenthal and R. Rosnow, Artifact in Behavioral Research. New York: Academic Press, 1969.

Rosenthal, Robert & K.L. Fode. "Three Experiments in Experimental Bias", Psychological Reports, 12: 491-511, 1963.

Rosenthal, Robert & Lenore Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart and Winston, 1968.

Rosenthal, Robert and Lenore Jacobson. "Self-fulfilling Prophecies in the Classroom: Teachers' Expectations as Unintended Determinants of Pupils' Intellectual Competence", In M. Deutsch, I. Katz and A. Jensen (eds.) Race, Social Class, and Psychological Development. New York: Holt, Rinehart and Winston, 1968.

Rosenthal, Robert and Lenore Jacobson. "Teachers' Expectancies: Determinants of Pupils' IQ Gains", Psychological Reports, 19: 115-118, 1966.

Rosenthal, Robert & R. Rosnow. Artifact in Behavioral Research. New York: Academic Press, 1969.

Rosenthal, Robert and D.B. Rubin. "Pygmalion Reaffirmed", Cambridge: Harvard University, 1971.

Roth, R.M. "Role of Self-Concept in Achievement", Journal of Experimental Education, 27: 265-281, 1959.

Rubovits, P.C. and M.C. Maehr. "Pygmalion Analyzed: Toward An Explanation of the Rosenthal-Jacobson Findings", Journal of Personality and Social Psychology, 19: 197-203, 1971.

Rushall, B.S. "Two Observational Scales for Sporting Environments", A paper presented at the First Canadian Congress for the Multi-Disciplinary Study of Sport and Physical Activity, Montreal, October, 1973.

Schrank, W. "A Further Study of the Labeling Effect of Ability Grouping", The Journal of Educational Research, 63: 358-360, 1970.

Snow, Richard E. "Unfinished Pygmalion", Contemporary Psychology, 14: 197-199, 1969.

Snygg, D. and A. Coombs. Individual Behavior. New York: Harper, 1949.

Stotland, E., S. Thorley, E.J. Thomas, and others. "The Effects of Group Expectations and Self-Esteem Upon Self-Evaluation", Journal of Abnormal and Social Psychology, 56: 223-229, 1958.

Strom, R.D. Psychology for the Classroom. New Jersey: Prentice-Hall, Inc., 1969.

Taylor, C. "The Expectations of Pygmalion's Creators", Educational Leadership, 28: 161-164, 1970.

Thorndike, R.L. "Review of 'Pygmalion in the Classroom' ", American Educational Research Journal, 5: 708-711, 1968.

Warner, W.L., R. J. Havighurst and M.B. Loch. Who Shall be Educated? New York: Harper and Brothers Publishers, 1944.

Webb, W.B. "A Procedure for Obtaining Self-Ratings and Group Ratings", Journal of Consulting Psychology, 20: 233-236, 1956.

Wesson, P.W. "A Preliminary Investigation into the Effects of Teacher Behavior in Elementary School Physical Education on the Self-Concept of Pupils", Unpublished master's thesis, University of Alberta, 1973.

Wright, C.J. and G. Nuthall. "Relationships Between Teacher Behaviors and Pupil Achievement in Three Experimental Elementary Science Lessons", American Educational Research Journal, 7: 477-492, 1970.

Yamomoto, K. (ed.) The Child and His Image: Self Concept in Early Years. New York: Houghton Mifflin, 1972.

APPENDIX A

INSTRUMENTS

NAME: _____

AGE: _____

(A) HOW WELL YOU THINK YOU ARE ABLE TO SWIM RIGHT NOW

I. Distance swim on your front

Excellent Very Good Good Fair Unsatisfactory

2. Distance swim on your back

Excellent Very Good Good Fair Unsatisfactory

3. Dive from deck

Excellent **Very Good** Good Fair Unsatisfactory

4. Treading water

Excellent Very Good Good Fair Unsatisfactory

5. Drownproofing

Excellent Very Good Good Fair Unsatisfactory

NAME: _____

(B) HOW WELL YOU THINK YOU WILL BE ABLE TO SWIM AT THE END OF THIS SWIMMING COURSE

1. Distance swim on your front

Excellent Very Good Good Fair Unsatisfactory

2. Distance swim on your back

Excellent Very Good Good Fair Unsatisfactory

3. Dive from deck

Excellent Very Good Good Fair Unsatisfactory

4. Treading water

Excellent Very Good Good Fair Unsatisfactory

5. Drownproofing

Excellent Very Good Good Fair Unsatisfactory

NAME:_____

(C) HOW WELL YOU THINK YOUR TEACHER WOULD RATE YOU RIGHT NOW

I. Distance swim on your front

Excellent Very Good Fair Unsatisfactory

2. Distance swim on your back

Excellent Very Good Fair Unsatisfactory

3. Dive from deck

Excellent Very Good Fair Unsatisfactory

4. Treading water

Excellent Very Good Fair Unsatisfactory

5. Drownproofing

Excellent Very Good Fair Unsatisfactory

STUDENT'S NAME _____

TEACHER RATING OF STUDENT'S SWIMMING ABILITY

Please rate the above named student according to how well you feel they are able to perform right now on each of the following items:

1. Distance swim on the front

   Excellent Very Good Good Fair Unsatisfactory

2. Distance swim on the back

   Excellent Very Good Good Fair Unsatisfactory

3. Dive from the deck

   Excellent Very Good Good Fair Unsatisfactory

4. Treading water

   Excellent Very Good Good Fair Unsatisfactory

5. Drownproofing

   Excellent Very Good Good Fair Unsatisfactory

SCORING SYSTEM

SWIMMING SKILLS TEST

Items	1 Point	2 Points	3 Points	4 Points	5 Points
Distance Swim On Front	Less than 1 length	Greater than 1, less than 2 lengths	Greater than 2, less than 3 lengths	Greater than 3, less than 4 lengths	Greater than 4 lengths
Distance Swim On Back	Less than 2 lengths	Greater than 2, less than 3 lengths	Greater than 3, less than 4 lengths	Greater than 4, less than 5 lengths	Greater than 5 lengths
Dive From Deck	Attempts, but jumps	Belly flop i.e stomach and hands hit simultaneously	Head first entry but sloppy, forward roll on entry	Clean head first entry. Too deep or too shallow	Very good dive. Good depth, good follow-through
Treading Water	Less than 1 min. 30 sec.	Greater than 1 min. 30 sec. Less than 3 min.	Greater than 3 min. Less than 4 min. 30 sec.	Greater than 4 min. 30 sec. Less than 6 min.	Greater than 6 min.
Drownproofing	Less than 1 min.	Greater than 1 min. Less than 2 min.	Greater than 2 min. Less than 4 min.	Greater than 3 min. Less than 4 min.	Greater than 4 min.

APPENDIX B

RAW DATA

RATINGS OF STUDENT'S SWIMMING ABILITY CONTROL GROUP

Subjects	Expert Rating		Self-Rating		Self-Rating (Teacher)		Self-Rating (Potential)	Teacher Rating
	Pre	Post	Pre	Post	Pre	Post		
1	15	18	22	15	23	21	25	17
2	6	19	10	22	11	21	23	10
3	14	12	15	18	13	17	18	17
4	6	7	16	16	18	17	19	9
5	13	10	16	19	17	20	20	19
6	9	8	17	17	25	16	23	8
7	6	19	11	14	12	14	17	18
8	7	23	19	23	20	23	23	24
9	10	14	10	18	9	15	15	15

RATINGS OF STUDENT'S SWIMMING ABILITY "SPECIAL POTENTIAL" GROUP

Subjects	Expert Ratings		Self-Rating		Self-Rating (Teacher)		Self Rating (Potential)	Teacher Rating
	Pre	Post	Pre	Post	Pre	Post		
1	12	11	16	20	16	19	22	16
2	6	9	13	15	11	15	16	7
3	8	8	8	19	12	17	14	9
4	14	17	18	21	17	21	23	19
5	7	11	19	22	16	21	21	18
6	5	12	19	18	20	17	24	9
7	13	16	16	17	13	17	14	14
8	12	10	15	15	13	15	13	17
9	6	6	14	17	17	19	20	7